文章编号：1673-9868(2012)11-0047-05

# 空气质量的混沌动力学分析①

张春涛，　刘学飞，　应　宏，　向瑞银

重庆三峡学院 数学与统计学院，重庆 404100

**摘要**：针对我国空气污染呈现的复杂非线性特征，提出一种新的空气质量的动力学分析方法．以三峡库区(万州)为研究地域．首先，以信息论为基础，对空气质量指数时间序列进行相空间重构，使用互信息法和条件熵法分别求解时间延迟和最小嵌入维数．然后，在重构的相空间上，计算空气污染指数时间序列的最大 Lyapunov 指数和关联维，研究空气污染的复杂性特征．研究结果表明，空气污染系统为混沌动力系统，这为空气污染的进一步研究提供了理论基础．

**关　键　词**：混沌；空气质量指数时间序列；相空间重构；最大 Lyapunov 指数；关联维

**中图分类号**：X51　　　　**文献标志码**：A

随着经济和社会的高速发展、城市化进程的加快，大气污染日益严重，空气质量问题已成为一个普遍关注的社会问题．空气污染系统是一个开放的、复杂的、非线性的大系统，同时又是一个动态的非平衡的复合系统，空气质量的变化过程受物理、化学、生物、气象以及人类活动等多种因素的影响，表现出貌似随机变化的复杂非线性特性．复杂系统在时间和空间上存在众多的状态变量，现有的基于数学表达式的分析方法由于很难将所有系统变量的关系都考虑进去，因此基于数学表达式的分析方法对复杂系统的处理具有一定的局限性．非线性科学理论方法可以从整体上揭示复杂系统的内在本质联系，已被引入到各种复杂问题的研究中[1-2]，因此可以采用非线性分析方法研究空气污染系统的时间演化过程[3-5]．

混沌理论是 20 世纪 80 年代发展起来的非线性科学，它的研究对象为非线性动力系统．混沌是非线性动态系统所特有的一种运动形式，它与时间序列分析方法相结合，便出现了混沌时间序列分析方法，混沌时间序列分析方法就是将一维时间序列扩展成高维相空间，从而使得相空间包括复杂系统主要的状态变量，然后在重构的高维相空间中研究系统状态变量的演化过程[6]．随着计算机科学和新兴交叉学科的迅猛发展，人们对混沌现象在自然科学领域和社会科学领域的表现有了更深刻的认识，混沌时间序列的分析方法在众多领域得到广泛的应用[6-7]．

空气质量变化时间序列可以视为空气污染系统一个状态变量的外在表现形式，该时间序列包含系统内其他状态变量信息，如何有效利用这些信息？借助 Takens 的嵌入定理[8]，时间序列重构的相空间在微分同胚下与原系统等价，因此可以在重构的相空间上研究空气质量系统各状态变量的动力学信息．本文首先建立了空气质量指数的基于信息论的相空间重构方法，使用互信息法和表征信息增量的条件熵法分别求解重构相空间的延迟时间和最小嵌入维数；然后在重构的相空间上对空气质量系统进行混沌动力学分析，计算其 Lyapunov 指数和关联维数，从而揭示空气质量演变的混沌动力学特征，为研究城市空气质量的演变分析提供新的方法．

① 收稿日期：2012-03-16
基金项目：重庆市教委科技资助项目(KJ111106)．
作者简介：张春涛(1978-)，男，重庆人，副教授，主要从事混沌动力系统的应用研究．

# 1 研究方法

在许多自然科学和工程技术等研究领域，人们比较容易获得的是研究对象的时间序列，传统的做法是直接从这个时间序列形式地分析系统的演化. 但由于时间序列是许多物理因素相互作用的结果，它蕴藏着参与运动的全部变量的信息，而直接的时间序列分析不能充分利用这些隐含信息，因此需将一维的时间序列扩展到高维相空间，这可使时间序列中隐含的系统信息充分地显露出来，此即是时间序列的相空间重构.

设混沌系统产生的时间序列为$\{x_i:i=1,2,\cdots,N\}$，根据嵌入定理，总存在合适的维数 $m$，时间延迟 $\tau$ 的相空间 $Y(i)=[x_i, x_{i+\tau}, \cdots x_{n+(m-1)\tau}]\in\mathbb{R}^m$，$i=1,2,\cdots,N-(m-1)\tau$，使得重构相空间中的“轨线”与原混沌系统在微分同胚意义下是等价的.

空气质量指数是评价空气污染的简单而又实用的指标[9]，因此本文使用空气质量指数时间序列研究空气污染系统的动力学特征. 本文以三峡库区腹地(万州区)为研究地域，选择 2008 年 6 月 5 日到 2012 年 1 月 31 日的空气质量指数日报时间序列为研究对象，具体数据如图 1 所示，从图 1 可以看出该时间序列表现出类似随机的演化过程.

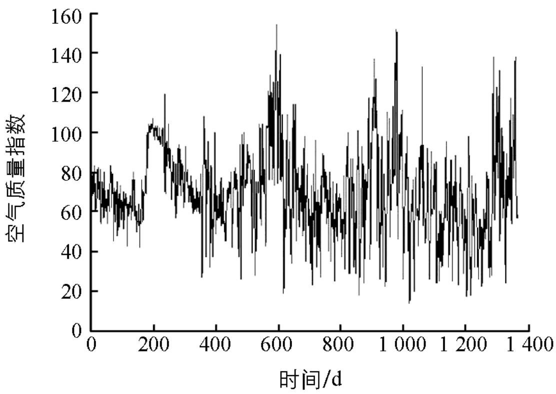


图 1 空气质量指数时间序列

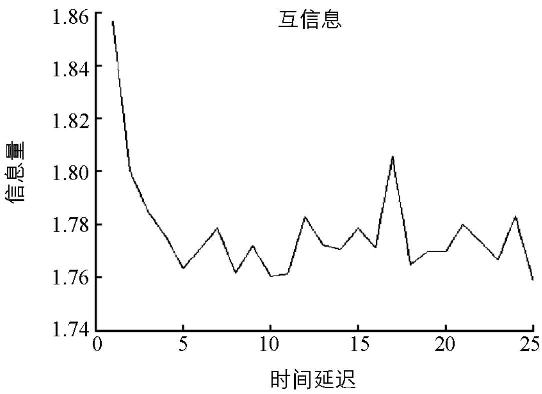


图 2 延迟时间图

重构过程的关键是确定两个重构参数：嵌入维数 $m$ 和时间延迟 $\tau$，这两个重构参数的信息论确定方法为互信息法和条件熵法.

## 1.1 延迟时间的选择——互信息法

互信息法[10]是计算重构相空间延迟时间的一种有效方法，它克服了自相关法[11]仅能提取序列间的线性相关性和较难应用于高维混沌系统的缺点.

对于系统 A 和系统 B，进行测量可能出现的结果是 $a_j$ 和 $b_k$. 设在 A 中观测到结果 $a_j$ 的概率为 $p_A(a_j)$，在 B 中观测到结果 $b_k$ 的概率为 $p_B(b_k)$，在 A 和 B 中观测到结果 $a_j$ 和 $b_k$ 的联合概率为 $p_{AB}(a_j, b_k)$，则互信息定义为：

$$I_{AB}=\sum_{a_j b_k} p_{AB}(a_j, b_k)\ln\frac{p_{AB}(a_j, b_k)}{p_A(a_j)p_B(b_k)} \tag{1}$$

互信息定量地刻画了系统 A 和 B 的相互关系，这种关系既包含线性关系又包含非线性关系，可以有效克服相关系数只反映线性关系的缺点.

对于时间序列 $x_i$ 和其延迟为 $\tau$ 的时间序列 $x_{i+\tau}$，设 $x_i$ 在时间序列中出现的概率为 $p(x_i)$，$x_{i+\tau}$ 在延迟时间序列中出现的概率为 $p(x_{i+\tau})$，$x_i$ 和 $x_{i+\tau}$ 在两个序列中共同出现的联合概率为 $p(x_i, x_{i+\tau})$. 联合概率 $p(x_i, x_{i+\tau})$ 可以通过在平面$(x_i, x_{i+\tau})$上数对应的格子得到，而格子一般采用等间距划分空间的方法得到，$p(x_i)$ 和 $p(x_{i+\tau})$ 即是联合概率 $p(x_i, x_{i+\tau})$ 的边缘概率. 于是，时间序列与其延迟时间序列的互信息就是延迟时间 $\tau$ 的函数：

$$I(\tau)=\sum p(x_i, x_{i+\tau})\log\frac{p(x_i, x_{i+\tau})}{p(x_i)p(x_{i+\tau})} \tag{2}$$

该函数度量了时间序列与延迟时间序列的统计依存关系，选择使 $I(\tau)$首次达到极小值的 $\tau$ 作为最佳延迟时

间，此时的时间序列与其延迟时间序列在充分保证两者之间独立性的同时又保留两者之间的动力学信息.

对选定的污染指数时间序列，用互信息法求解延迟时间(图 2). 从图 2 可以看出，在 $\tau=5$ 时，$I(\tau)$ 第一次达到极小值，因此空气质量指数时间序列的最佳延迟时间选定为 $\tau=5$.

### 1.2　嵌入维数的选择——条件熵法

对非线性动力学系统，各状态变量是相互作用、相互关联的，系统的任一变量演化都包含其他变量的演化信息，根据 Takens 嵌入定理，通过重构可以从一变量时间序列中提取原系统动力学特征. 另一方面，相空间重构就是要使得重构的相空间包含足够多的原系统信息，并且要避免冗余信息的产生. 本文采用重构坐标的条件熵表示不同维数下的信息增量确定最小嵌入维数.

设延迟时间为 $\tau$，$d$ 维的相空间为 $[x_i, x_{i+\tau}, \cdots, x_{i+(d-1)\tau}]$，它所具有的信息量为

$$H_d \triangleq H(x_i, x_{i+\tau}, \cdots, x_{i+(d-1)\tau}) \tag{3}$$

当相空间的维数增加到 $d+1$ 维时，相空间的信息量为

$$H_{d+1} \triangleq H(x_i, x_{i+\tau}, \cdots, x_{i+d\tau}) \tag{4}$$

由维数增加 1 所引起的信息量的变化为条件熵：

$$H_{d+1\mid d} \triangleq H(x_{i+d\tau} \mid x_i, x_{i+\tau}, \cdots, x_{i+(d-1)\tau}) = H_{d+1} - H_d \tag{5}$$

可见如果 $d$ 维的相空间包含了足够多的系统信息，条件熵 $H_{d+1\mid d}$ 不随维数 $d$ 的增加而变化，即 $H_d=H_{d+1}$，此时 $d$ 为一个合适的嵌入维数，如果 $H_{d+1}\gg H_d$，可以认为低维的重构相空间没有完全包含系统信息，此时的 $d$ 不是一个合适的嵌入维数.

如果时间序列是确定的，则嵌入维数是存在的，随着嵌入维数 $d$ 的增加，条件熵 $H_{d\mid d+1}$ 的值逐渐变小，且 $H_{d\mid d+1}$ 将在 $d$ 大于某一特定值 $d_0$ 时不再变化. 因此可以根据 $H_{d\mid d+1}$ 是在急速变化还是缓慢变化并且趋于稳定来确定最小嵌入维数，在实际应用中，取 $H_{d\mid d+1}$ 变化的一个明显转折并且其后趋于平稳的 $d$ 为最小嵌入维数. 本文采用矩阵标识法[12]计算高维空间的联合熵.

对选定的时间序列，用条件熵法求解最小嵌入维数，条件熵随嵌入维数变化曲线如图 3 所示. 从图 3 可以看出，当维数增加到 3 时条件熵曲线出现一个转折，并且其后的条件熵趋于平稳，因此对选定的污染指数时间序列取最小嵌入维数 $m=3$.

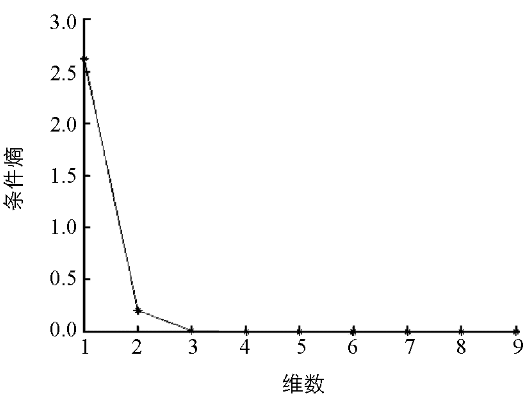


**图 3　嵌入维数图**

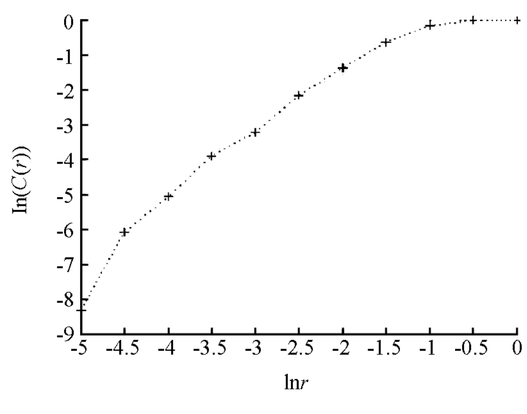


**图 4　lnC(r)－lnr 关系曲线**

## 2　空气质量系统的混沌动力学分析

确定好重构相空间的两个重构参数后，选定的空气质量指数时间序列重构的相空间为 $[x_i, x_{i+5}, x_{i+10}]\in\mathbb{R}^3$，由于重构的相空间等价于空气质量系统，所以我们就在重构的相空间上讨论原系统的动力学特征，下面主要研究空气质量指数的最大 Lyapunov 指数和关联维的计算.

### 2.1　最大 Lyapunov 指数的数值计算方法

Lyapunov 指数是根据相轨迹有无扩散运动来判定系统的混沌特性. Lyapunov 指数大于零表明相轨道局部分离，系统表现出奇怪吸引子，因此只要有一个 Lyapunov 指数大于零，就可以说该系统具有混沌特性. 在实际中通常只需估计最大 Lyapunov 指数 $\lambda_1$，目前计算最大 Lyapunov 指数的方法很多，本文选择

Wolf 方法[13]来计算最大 Lyapunov 指数.

对选定的空气质量指数时间序列 $\{x_i\}$，根据信息论方法确定的重构参数：$m=3$，$\tau=5$，则重构相空间的相点为：

$$\begin{cases} Y(1)=(x_1,\ x_6,\ x_{11}) \\ Y(2)=(x_2,\ x_7,\ x_{12}) \\ \quad\vdots \\ Y(i)=(x_i,\ x_{i+5},\ x_{i+10}) \\ \quad\vdots \\ Y(n)=(x_n,\ x_{n+5},\ x_{n+10}) \end{cases} \tag{6}$$

式中：$Y(i)$ 为重构相空间的第 $i$ 个相点，$n=N-(m-1)\tau$ 表示相空间相点的个数.

取初始相点为 $Y(t_0)$，设其与最近邻点 $Y_0(t_0)$ 的距离为 $d_0(t_0)$，追踪这两点的时间演化，直到在某 $t_1$ 时刻其间距超过某规定值 $\varepsilon$，其距离演化为 $d_1(t_1)$，保留 $Y(t_1)$，并找 $Y(t_1)$ 的另一邻近点 $Y_1(t_1)$，使得 $d_0(t_1)=\|Y(t_1)-Y_1(t_1)\|<\varepsilon$，并且 $Y_0(t_1)-Y(t_1)$ 与 $Y_1(t_1)-Y(t_1)$ 之间的夹角尽可能小，继续上述过程，直到 $Y(t)$ 到达时间序列构成的相空间的终点，则最大 Lyapunov 指数为：

$$\lambda_1=\frac{1}{t_M-t_0}\sum_{k=1}^{M}\log\frac{d_1(t_k)}{d_0(t_{k-1})} \tag{7}$$

其中 $M$ 为追踪演化过程总的迭代次数.

通过 Matlab 编程实现上述算法，计算得空气质量指数时间序列的最大 Lyapunov 指数为 $\lambda_1=0.253\,7$. $\lambda_1>0$，说明空气污染具有混沌特性.

### 2.2 关联维的计算

关联维是另一个重要的混沌特征量，本文采用 G-P 算法[14]计算关联维数，其计算步骤为：

a. 首先对时间序列进行归一化处理；

b. 设定欧氏距离极限 $r$ 的取值范围；

c. 对归一化后的时间序列 $x_i$ 进行相空间重构，即相点 $Y(i)=(x_i,\ x_{i+\tau},\ \cdots,\ x_{i+(m-1)\tau})$；

d. 从某参考点 $Y(i)$ 开始，计算其余相点 $Y(j)$ 与 $Y(i)$ 间的欧氏距离，在距离极限 $r$ 范围内改变 $r$ 的取值，并计算不同 $r$ 下关联积分：

$$C(r)=\frac{1}{n(n-1)}\sum_{i,\ j=1,\ i\neq j}^{n}H(r-\|Y(i)-Y(j)\|) \tag{8}$$

其中：$H(x)$是 Heaviside 函数，当 $x\leqslant 0$ 时 $H(x)=1$，当 $x>0$ 时 $H(x)=0$；$n=N-(m-1)\tau$ 为相空间点数.

e. 用最小二乘法计算 $\ln C(r)$ 和 $\ln r$ 双对数曲线的斜率即为关联维数 $D$.

根据上面的计算方法，通过 Matlab 编程得 $\ln C(r)-\ln r$ 的关系图如图 4 所示，然后用最小二乘法拟合数据计算空气质量指数时间序列的关联维数 $D=1.770\,7$，关联维数为分数，说明空气污染的吸引子具有分数维. 同时说明了最小嵌入维数选择和合理性，因为实际选择的最小嵌入维数 $m$ 一般应满足：$D+1<m<2D+1$.

## 3 总　结

本文在基于信息论的相空间重构方法基础上对三峡库区腹地(万州)的空气质量指数时间序列进行分析研究，通过相空间重构反演复杂的非线性的多变量的空气污染系统，从总体上把握空气污染的复杂特征，发现空气污染具有混沌特性. 应用表明，基于信息论的相空间重构方法的有效性，同时相空间重构技术很好地揭示空气质量系统的混沌特征量：最大 Lyapunov 指数大于零，吸引子维数为分数，正是这种混沌特性，使得空气质量变化曲线呈现出一定的不规则、貌似随机的特点. 该研究充分利用了环境监测数据，避免了机理分析所需要的苛刻条件，可以很好地为城市大气污染的防治提供决策支持.

**参考文献：**

[1] 张　丽，杨庆媛，冯应斌. 基于分形理论的区域土地利用类型探讨 [J]. 西南大学学报：自然科学版，2008，30(8)：137－141.

[2] 胥昌纯，卢　挺，陈刚才. 利用复杂网络理论研究三峡库区流域水华爆发 [J]. 西南师范大学学报：自然科学版，2011，36(3)：240－244.

[3] RAGA G B，LE M. On the Nature of Air Pollution Dynamics in Mexico City-I. Nonlinear Analysis [J]. Atmospheric Environment，1996，30：3987－3993.

[4] WINDSOR H L，TOUMI R. Scaling and Persistence of UK Pollution [J]. Atmospheric Environment，2001，35：4545－4556.

[5] 史　凯，刘春琼，艾南山. 上海市空气质量变化的多重分形分析 [J]. 环境污染与防治，2008，30(9)：60－64.

[6] 吕金虎，陆君安，陈士华. 混沌时间序列分析及其应用 [M]. 武汉：武汉大学出版社，2002.

[7] 陆仁强，牛志广，张宏伟. 基于相空间重构理论的天津市近海水质混沌特性研究 [J]. 海洋环境科学，2010，29(1)：104－107.

[8] TAKENS F. Dynamical Systems and Turbulence，Lecture Notes in Mathematics [M]. Berlin：Springer-Verlag，1981：366－381.

[9] 李　杰，范　毅，曾雪梅. 重庆市南岸区环境空气质量现状及变化趋势研究 [J]. 西南师范大学学报：自然科学版，2010，35(4)：184－189.

[10] FRASER A M，Swinney H L. Independent Coordinates for Strange Attractors form Mutual Information [J]. Physical Review A，1986，33：1134－1140.

[11] KANTZ H，SCHREIBER T. Nonlinear Time Series Analysis [M]. Cambridge：Cambridge University Press，2004.

[12] 肖方红，阎桂荣，韩宇航. 混沌时序相空间重构参数确定的信息论方法 [J]. 物理学报，2005，54(2)：550－556.

[13] WOLF A，SWIFT J B，Swinney H L，et al. Determining Lyapunov Exponents from Time Series [J]. Physica D，1985，16(2)：285－371.

[14] GRASSBERGER P，PROCACCIA I. Measuring the Strangenss of Strange Attractors [J]. Physica D，1983，9：189－208.

# Chaotic Dynamic Analysis of Air Quality

ZHANG Chun-tao，　LIU Xue-fei，　YING　Hong，　XIANG Rui-yin

*College of Mathematics and Statistics，Chongqing Three Gorges University，Chongqing 404100，China*

**Abstract**：Taking into consideration the fact that the distribution of air pollution in China is characterized by complexity and non-linearity，this paper proposes a new approach for analyzing the dynamic characteristics of air quality. Taking the Three Gorges Reservoir Area (Wanzhou) as the research object and based on the information theory，it uses the phase space reconstruction method to reconstruct the phase space for air quality index time series and employs the mutual information method and the conditional entropy method to determine the time delay and the minimum embedding dimension，respectively. Then，it studies the complex characteristics of air pollution through calculating the biggest Lyapunov exponents and attractor dimension of the air pollution index time series in the reconstructed phase space. The research results show that air pollution is a chaotic chaos dynamic system.

**Key words**：chaos；air quality index time series；phase space reconstruction；biggest Lyapunov exponent；correction dimension

**责任编辑**　潘春燕